<u>REMARKS</u>

The present application was filed on October 31, 2000 with claims 1-27. Claims 1, 10 and 19 are the independent claims.

In the outstanding Office Action, the Examiner: (i) rejects claims 1-27 based on 35 U.S.C. §112, first and second paragraphs; and (ii) maintains the rejection of claims 1-27 under 35 U.S.C. §102(e) as being unpatentable over U.S. Patent No. 6,112,203 to Bharat et al. (hereinafter "Bharat").

In this response, Applicants amend independent claims 1, 10 and 19, and traverse the §112 and §102(e) rejections for at least the following reasons.

Regarding the §112, first paragraph and second paragraph, rejections of claims 1-27, while Applicants believe that the previous amendment and specification more than sufficiently satisfy both paragraphs, Applicants have nonetheless further defined the claimed invention in a sincere effort to expedite the case through to issuance.

Independent claim 1 now recites a computer-based method of performing document retrieval in accordance with an information network, the method comprising the steps of: initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without use of an initial link structure; collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed; and using the collected statistical information to automatically determine further document retrieval operations, wherein the statistical information using step further comprises learning a link structure from at least a portion of the collected statistical information with each successive document retrieval operation. Independent claims 10 and 19 recite similar limitations.

The Office Action suggests that the previously-added claim 1 language was "inoperative and impossible" and that the specification does not reasonable support the limitation. While Applicants do not agree, they have further clarified the limitation.

Support for the amendment may be found throughout the present specification. As illustratively explained in the present specification at page 4, line 22, through page 5, line 20:

> The present invention provides a more interesting and significantly more general alternative to conventional crawling techniques. As is evident from the teachings herein, <u>no</u>

7

<u>specific model for web linkage structure is assumed in intelligent crawling according to the invention. Rather, the crawler gradually learns the linkage structure statistically as it progresses.</u> By linkage structure, we refer to the fact that there is a certain relationship between the content of a web page and the candidates that it links to. For example, a web page containing the word "Edmund Guide" is likely to link to web pages on automobile dealers. In general, linkage structure refers to the relationship between the various features of a web page such as content, tokens in Universal Resource Locators (URL), etc. Further, in general, it is preferred that the linkage structure be predicate-dependent. An intelligent crawler according to the invention learns about the linking structure during the crawl and find the most relevant pages. Initially, the crawler behavior is as random as a general crawler but it then gradually starts auto-focusing as it encounters documents which satisfy the predicate. A certain level of supervision in terms of documents which satisfy the predicate may be preferred since it would be very helpful in speeding up the process (especially for very specific predicates), but is not essential for the framework of the invention. This predicate may be a decision predicate or a quantitative predicate which assigns a certain level of priority to the search.

The intelligent crawler of the invention may preferably be implemented as a graph search algorithm which works by treating web pages as nodes and links as edges. The crawler keeps track of the nodes which it has already visited, and for each node, it decides the priority in which it visits based on its understanding of which nodes is likely to satisfy the predicate. <u>Thus, at each point the crawler maintains candidate nodes which it is likely to crawl and keeps re-adjusting the priority of these nodes as its information about linkage structure increases</u> (Underlining added for emphasis).

As further support, the present specification, starting at page 8, line 22, explains that the input to the intelligent web crawling process of FIG. 2 includes a list of URLs to web pages from which the crawl starts, and a predicate which is used to focus and control the crawl. That is, there is no link structure (e.g., as explained above, a structure representing the relationship between the various features of a web page) that is assumed in the initial step of the retrieval operation. A link structure is then gradually learned as the process progresses, as further illustrated in FIG. 2 and the subsequent figures.

In view of the above, Applicants respectfully request withdrawal of the §112, first paragraph and second paragraph, rejections of claims 1-27.

Regarding the §102 rejection of claims 1-27 based on Bharat, Applicants respectfully assert that Bharat fails to teach or suggest all of the limitations of claims 1-27.

While Applicants believe that the claims of the present application in their form prior to the previous amendment, dated September 26, 2005, were patentably distinguishable over Bharat for

at least the reasons given in their Appeal Brief dated March 28, 2005, Applicants have nonetheless amended the independent claims again in a sincere effort to expedite the present application through to issuance.

Independent claim 1 now recites a computer-based method of performing document retrieval in accordance with an information network, the method comprising the steps of: initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without use of an initial link structure; collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed; and using the collected statistical information to automatically determine further document retrieval operations, wherein the statistical information using step further comprises learning a link structure from at least a portion of the collected statistical information with each successive document retrieval operation (underlining is to emphasize the added language).

In contrast, Bharat discloses a method for ranking documents in a hyperlinked environment using connectivity and content analysis (see Abstract of Bharat). Thus, Bharat does not teach or suggest that "an initial document retrieval operation is performed without use of an initial link structure" and that "a statistical information using step further comprises learning a link structure from at least a portion of the collected statistical information with each successive document retrieval operation," as recited in the claimed invention.

As is made clear at columns 2 and 3, the techniques of Bharat require a preconstructed graph that includes nodes and directed edges, where each node represents a document and the directed edges represent the links connecting the documents. This "start-set" of Bharat is a link structure. Thus, unlike the claimed invention, Bharat assumes an initial link structure.

However, while an illustrative method of the invention may generate a graph-like (linkage) structure as it "crawls," it merely starts off with a start list which is merely a list of Uniform Resource Locators (URLs), see, for example, page 8 and 9 of the present specification. As explained above, the method keeps track of the nodes which it has already visited, and for each node, it decides the priority in which it visits based on its understanding of which nodes is likely to satisfy the predicate. Thus, at each point, the method maintains candidate nodes which it is likely

to crawl and keeps re-adjusting the priority of these nodes as its information about linkage structure increases. This is an illustrative example of what is meant by "an initial document retrieval operation is performed without use of an initial link structure" and that "a statistical information using step further comprises learning a link structure from at least a portion of the collected statistical information with each successive document retrieval operation," as recited in the claimed invention. Bharat does not teach or suggest this limitation.

Applicants assert that there is clearly a significant difference between the start-set of Bharat, which requires generating or obtaining an initial preconstructed graph or link structure, and the start list illustratively disclosed in the present specification, which merely contains a list of URLs that are used to perform the initial retrieval operation. Generation of the list of URLs does not require preconstruction of an express linkage structure between the URLs. The start list of the illustrative invention and the start set of Bharat are two different inputs.

Applicants also assert that Bharat does not disclose the step of "collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed," as in the claimed invention. While Bharat does disclose content analysis, it does not appear that any "statistical information" is being collected in the Bharat document ranking technique. Therefore, whatever definition the Office Action suggests for "statistical information," Bharat still does not collect statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed.
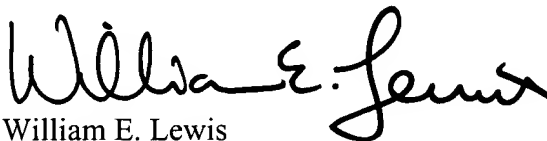
For at least the above reasons, Applicants respectfully assert that independent claims 1, 10 and 19 are patentable over Bharat.

The remainder of the claims (namely, claims 2-9, 11-18 and 20-27) rejected over Bharat depend, either directly or indirectly, from claims 1, 10 or 19, which are believed patentable for the reasons set forth above. Furthermore, the remaining claims define additional patentable subject matter in their own right.

In view of the above, Applicants respectfully request withdrawal of the §102(e) rejections of claims 1-27.

Applicants believe that claims 1-27 are in condition for allowance, and respectfully request favorable reconsideration.

Respectfully submitted,

Date: March 23, 2006

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946